PCT/AU98/00769

AUSTRALIA

09/508713 REC'D 02 OCT 1998 WIPO PCT

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patent Office
Canberra

I, KIM MARSHALL, MANAGER EXAMINATION SUPPORT AND SALES, hereby certify that the annexed is a true copy of the Provisional specification in connection with Application No. PP 2714 for a patent by LAKE DSP PTY LTD filed on 31 March 1998.

I further certify that the annexed specification is not, as yet, open to public inspection.

COMMONWEALTH OF AUSTRALIA
AUSTRALIA
PATENT OFFICE

WITNESS my hand this Twenty Fourth day of September 1998

KIM MARSHALL
MANAGER EXAMINATION SUPPORT AND SALES

P/00/009
Regulation 3.2

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

Application Title: Low Memory and Computation Filtering Effects in Spatialization of Stereo Headphone Devices

The invention is described in the following statement:

GH REF: **PP24042-AA/PJT**

A Low Memory and Computation Filtering Effects in Spatialization of Stereo Headphones (PAT48)

Field of the Invention

The present invention relates to the creation of sound environments around a listener and, in particular, where the listener is listening to the sound environment via headphones.

Background of the Invention

A number of different sound reproduction techniques are in popular use. These techniques are created so as to provide a volumetric rendering of a sound such that it takes on spatial components. Historically, most sound was initially produced in a "mono" signal format. At present, however, one of the most popular formats is a stereo format wherein two sound signals are produced or transmitted such that, when output on a pair of speakers, they appear to have a spatial component or environment out of the front of a listener when those speakers are placed in front of the listener.

Unfortunately, when standard headphones are utilised, the out-of-head perception is lost and the sound appears to be coming from somewhere inside the listeners head and is substantially centralized.

Other sound formats face similar problems when reproduced over headphones. For example, the Dolby AC-3 format, another popular format, is designed for the placement of a number of speakers around a listener so as to create a substantially richer sound environment. Again, when headphone devices are utilised in such an environment the intended spatial location of the sound is lost and again the sound appears to come from within the head of a listener.

Summary of the Invention

It is an objection of the present invention to provide an improved method and system which allows for the playback of audio through headphones so as to create the illusion of

sound sources external to the listener's cranium. The system includes improvements which relate to the reduction in computational requirements of existing systems and improving the realism of a virtual speaker systems. The system provides for the production of a stable illusion of sound sources positioned around the user with an impression of a depth and distance and thereby provides a richer environment for the headphone listener.

In accordance with a first aspect of the present invention, there is provided an apparatus for creating, utilizing a pair of oppositely opposed headphone speakers, the sensation of a sound source being spatially distant from the area between the pair of headphones, the apparatus comprising: (a) a series of audio inputs representing audio signals being projected from an idealized sound source located at a spatial location relative to the idealised listener; (b) a first mixing matrix means interconnected to the audio inputs and a series of feedback inputs for outputting a predetermined combination of the audio inputs as intermediate output signals; (c) a filter system of filtering the intermediate output signals and outputting filtered intermediate output signals and the series of feedback inputs, the filter system including separate filters for filtering the direct response and short time response and an approximation to the reverberant response, in addition to feedback response filtering for producing the feedback inputs; and (d) a second matrix mixing means combining the filtered intermediate output signals to produce left and right channel stereo outputs.

Preferably, a predetermined number of the feedback inputs are also input to the second matrix mixing means. The feedback response filtering can comprise a reverberation filter. The reverberation filter can comprise one of a sparse tap FIR, a recursive algorithmic filter or a full convolution FIR filter and the audio inputs can comprise a surround sound set of signals.

Further, in one embodiment the feedback inputs are mixed with the frontal portions of the audio inputs only.

The filter system can include a front sum filter filtering a summation of the audio inputs positioned in front of the idealized listener and the front sum filter comprises substantially an approximation of the sum of a direct and shadowed head related transfer function for the front inputs. Further, the filter system can include a front difference filter filtering a difference of the audio inputs positioned in front of the idealized listener and the front difference filter comprises substantially an approximation of the difference of a direct and shadowed head related transfer function for the front inputs. Further, the filter system can include a rear sum filter filtering a summation of the audio inputs positioned in rear of the idealized listener and the rear sum filter comprises substantially an approximation of the sum of a direct and shadowed head related transfer function for the rear inputs. Further, the filter system can include a rear difference filter filtering a difference of the audio inputs positioned in rear of the idealized listener and the rear difference filter comprises substantially an approximation of the difference of a direct and shadowed head related transfer function for the rear inputs. Further, the filter system can include a reverberation filter interconnected to the sum of the audio inputs.

Brief Description of the Drawings

Notwithstanding any other forms which may fall within the scope of the present invention, preferred forms of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Fig. 1 illustrates the operation of a system of the present invention;

Fig. 2 illustrates a generalised form of the preferred embodiment;

Fig. 3 illustrates a more detailed schematic form of

the preferred embodiment;

Fig. 4 illustrates a schematic diagram of a Dolby AC-3 to stereo headphone converter;

Fig. 5 illustrates a stereo input to stereo output embodiment in schematic form;

Fig. 6 illustrates in schematic form, one form of conversion from Dolby AC-3 inputs to stereo outputs in accordance with the present invention;

Fig. 7 illustrates a modified general embodiment;

Fig. 8 illustrates a schematic diagram of a modified form of stereo mixing;

Fig. 9 illustrates a modified form of surround sound mixing;

Fig. 10 illustrates the process of calculation of direct and shadowed responses;

Figs. 11 and 12 illustrate resultant direct and shadowed responses;

Fig. 13 illustrates a suitable reverb sparse tap;

Figs. 14 and 15 illustrate suitable reverb filters.

<u>Description of Preferred and Other Embodiments</u>

A number of the embodiments of the present invention will be described for different sound formats.

Turning initially to Fig. 1, there is provided a schematic illustration of the operation of a first embodiment of the invention. In this embodiment, a series of audio inputs 11 are provided to a mechanism 12 which would normally form part of the prior art taking the audio signal inputs and creating a series of speaker feeds 13. The speaker feeds 13 can be provided for the various output formats, for example stereo output formats or AC-3 output formats. The operation of the portion within dotted line 14 being entirely conventional. The speaker feeds are forwarded to the headphone processing system 15 which outputs to a set of standard headphones 16 so as to simulate the presence of a number of speakers around the listener

using headphones 16.

Fig. 1 illustrates the example where headphone processing system 16 simulates the presence of two virtual speakers 17, 18 in front of the user of headphones 16 as would be the normal stereo response. The arrangement of Fig. 1 has particular advantages in that it can be incorporated in any system that is generally utilised for the playback of stereo audio. The system processes the usual signals intended for playback over speakers and is therefore compatible with and can be used in conjunction with any other system designed for enhancing the reproduction of audio over loudspeakers.

The general structure of a first form and implementation of headphone processing system is by a filter structure where each of the intended speaker feeds is passed through two filters, one for each ear. The resultant sum of all these filters is the signal sent to the appropriate headphone channel for that ear. In alternative embodiments, the filters may or may not be updated to reflect changes in the orientation of the listener's head inside the virtual speaker array. By updating the filters based on the physical orientation of a listener's head, a more imersive head-tracked environment can be created. Various implementations can be variations on this theme so as to reduce computational requirements. Further, non-linear, active or adaptive components can be added to the structure to improve performance.

An example of the general structure a headphone processing system is in a more complex form is illustrated in Fig. 2. The implementation 20 includes a series of speaker feeds e.g. 21 each of which has a separate filter e.g. 22, 23 applied with one filter 22 being applied for a left hand channel and one filter 23 being applied for a right hand channel. The filter outputs are summed e.g. 24 together to form a final output 25.

The arrangement of Fig. 2 can lead to overburdening

complexity in a large number of filters e.g. 22 must be provided which is likely to substantially increase costs. A technique for significantly reducing the computational requirements by taking advantage of symmetry is to utilise "shuffling" techniques. For a pair of channels, this represents applying filters to the sum and difference of the channels before recombination. For the stereo case where the filters are symmetric (i.e. FilterLL = FilterRR, FilterLR = FilterRL) this can reduce the computational requirements by 50%. This technique can be represented by inserting a linear matrix mix before and after the filter banks.

More generally, as indicated in Fig. 3, the implementation structure 30 can consists of:

* A number of inputs 31
* A mixing matrix 32 to produce a set of signals each of which is a linear combination of the input signals (note the intermediate set of signals may include the input signals themselves and may include duplicate signals). In alternative embodiments, the matrix gains may be time varying.
* A series of filters e.g. 33 on each of the intermediate signals. The filters can be independent and thus can have different structures, lengths and delays (for example IIR, FIR, sparse tap IR, and low latency convolution).
* A mixing matrix 35 to combine the filtered intermediate signals appropriately to create the two headphone output signals 36.

Some specific implementations of the general system of Fig. 3 are as follows:

High End AC-3 Decoder

As illustrated in Fig. 4, the Dolby™ AC-3 standard defines a set of 5 (.1) channels to be used as speaker feeds 41. These channels are derived from an AC-3 bit stream data source using an AC-3 decoder. Once decoded, the speaker

feeds are suitable for utilisation as inputs 41 to the arrangement 40 of Fig. 4 which produces headphone outputs 42. Each of the five speaker feeds is passed through a filter e.g. 43, 44 for each ear and summed e.g. 45 to produce the headphone signal - making a total of 10 filters.

The filters are provided to simulate a corresponding virtual speaker array.

To achieve a high level of quality in the simulation of a virtual speaker array, fairly long filters are required to take into account the spatial geometry of the listening environment. With proper filter sets (incorporating equalisation for the headphones and proper head related transfer functions) the results provide close to a perfect illusion of a set of external speakers being used.

The 10-filter design can be refined to reduce computational power without too much quality degradation by using 10 shorter filters and only two full-length filters. The two longer filters 47, 48 can be a binaural simulation of the tail of an average room response. A combination of all 5 speaker feeds is fed via summer 49 into the binaural tail filters 47, 48 to give an approximation of the real room response. Each of the short filters e.g. 43, 44 can be the early part of the response for that particular speaker to the listener's ear.

The filter length used in prototype implementations can be typically 2000 taps at 48kHz sampling rate for the short filters e.g. 43, 44 and 32000 taps for the longer filters 47, 48. The long filters usually have a lower bandwidth and can be implemented with latency - this can be taken advantage of using a reduced sample rate processing to lower the computational requirements. The filters can be implemented using low latency convolution algorithms to lower the system latency and computational requirements.

The filter sets can be obtained by simulating a virtual speaker set-up using acoustic modelling packages such as CATT acoustics or by using a real or synthetic head placed

inside a real speaker array.

The High End AC-3 decoder 40 provides a fairly accurate simulation through headphones of a virtual speaker array, however, it also includes a large amount of computational resource.

Low End Stereo Decoder

The Low-End Stereo Decoder as illustrated 50 in Fig. 5, is a device utilising only some of the features of the high-end computationally resourced system. The main aim is to manipulate a stereo source 51 for playback over headphones 52 to give the impression of the sound originating from around the listener, simulating the experience of listening to a well configured stereo. The system of Fig. 5 is designed to be suitable for mass production at a low cost; thus the more important issues of the design are in reducing the computational complexity.

As noted previously, the general structure of the low-end stereo decoder 50 has two inputs 51 for conventional stereo and two outputs 52 for the headphone signals. A bank of two filters is used, operating on the sum 55 and difference 56 signals of the input stereo pair 51.

The low end stereo decoder 50 is another example, consistent with the general implementation outlines previously. In this case the matrix operations are a two channel sum 55 and difference 56 shuffle. The filters are applied to the sum and difference signals to half the computational requirements where the desired result is symmetric (i.e. L->L=R->R and L->R=R->L).

The performance of this system is dependent on the choice of filter coefficients. To reduce the computational requirements, short filters are ideally used. It has been found that the difference filter can be somewhat shorter than the sum filter and still produce a reasonable result.

The preferred form is to use a set of filters that is a combination of the head related transfer functions for 30° in the horizontal plane, and a semi-reverberant tail but

fairly sparse filter. The filter construction can be as follows:

Given the following impulse responses

D Direct ear response - normalised to unity energy

S Shadowed ear response - scaled in proportion to D

R Reverberant response - normalised to unity energy

and the following parameter

α Presence - the amount of reverberant feed in the mix

then the following filters are applied to the sum and difference signals to produce new Sum′ and Diff′ signals

$$Sum' = \left(\sqrt{(1-\alpha^2)}(D+S)+\alpha R\right) \otimes Sum$$

$$Diff' = \left(\sqrt{(1-\alpha^2)}(D-S)\right) \otimes Diff$$

To further reduce the amount of processing required a number of approximations can be made to the filter set. The direct ear response is assumed to be unity. The shadowed ear response can be approximated by a 5 tap FIR matching the frequency response and group delay of the exact signal derived from deconvolving a direct ear response from the appropriate shadowed response. Around 20 sparse taps can approximate the reverberant response from a 5-10ms delay line.

With this approach it has been found that the coefficients can be heavily quantised and reasonable performance maintained. The sum filter can be implemented as a set of 25 taps from a 256 tap delay line (at 48kHz) while the difference filter can be mere 6 taps from a 30 tap delay line. This allows the system to be implemented using around 3 MIPS thus making it suitable for low cost, mass production and incorporation into other audio products using headphones.

Further extensions to the implementation 50 can include:

* The use of low-latency convolution to allow the possibility of longer filters.

* The addition of further inputs and similar budget processing to allow for the simulation of "surround sound" formats. For example, a surround channel could be added that simulates the presence of sounds behind or around the rear of the listener.

* Incorporation of budget head tracking processing to change the early HRTF components to give a sense of stationary sound sources when the head is rotated.

* Addition of non-symmetric components to provide better performance when the stereo signal has significant mono components in the mix.

* Addition of non-linear components to enhance the performance (for example a dynamic range compressor to improve the quality of listening in a noisy environment).

It can therefore be seen that the first series of embodiments utilise a unique combination of input mix-processing, filters and output mix-processing to create the appearance of 3-dimensional sound over headphones. The arrangements disclosed include reduced computational complexity and memory requirements resulting in a significant reduction in implementation costs. The filter structures and coefficients improve the directionality and depth of the sound with minimal increase in computational complexity. The simple HRTF approximations require little processing power having been significantly reduced from the normal 50-60 filter taps.

The significant HRTF features include

a) the significant main energy component of the direct response (short time approximation) and the approximation of the convolution mapping of the direct response to the shadow or reflected response.

(b) the use of filter coefficients comprising a 5-10ms sparse tap filter after about 50-100 taps. The use of the reverberant filter enhances the performance of the HRTF

approximations, normal HRTF's and room impulse responses by increasing the localisation and depth of sound.

(c) In a modification, the HRTF approximations can include coefficients for containing anti-phase component in the shadow response so as to improve rear localisation.

(d) The filters of preferred embodiments include a first part which provides directionality and localisation and a second part which provides ambiance and room acoustics but minimal directionality.

The utilisation of the delivery format of the preferred embodiments provides considerable flexibility in the trade off of optimal computation and memory usage versus performance.

The extension of the system 50 of Fig. 5 to Dolby AC-3 inputs can be as shown 60 in Fig. 6. The center channel 61 is added 62, 63 to the front left and rear right channels respectively. The output signals are fed to delay units 64, 65 which can be 5 to 10 msec delay lines, before being fed to HRTFs 67 - 69 which provide outputs for summing 70, 71 to the left and right ears. The rear signals 73, 74 are used to form sum and difference signals 76,77 which are fed to HRTFs 79, 80 which provide anti-phase to the summing units 70, 71.

Turning now to Fig. 7 there is illustrated a modified form of general structure 90 silicone .... Fig. 3. However, the arrangement of Fig. 7 includes filters 91, 92 and feedback path 93. The mixing matrix 94 remains a simple linear matrix with the ability to negate, scale, sum and redirected its input signals as required for a specific implementation. The outputs 93 of the feedback filters 91, 92 also go into a second mixing matrix (not shown) in a alternative embodiment, to contribute directly to the output. In an even more general arrangement, all filter outputs can be fed back to the first mixing matrix 94 at which point there may be included or excluded from the mix. However, generally it is preferably to keep the size of the

mixing matrix to a minimum.

The modified general structure 90 allows for a feedback path 93 other than a recursive element within each separate filter. A more realistic reverberation can be created by feeding the outputs of a reverb filter created as part of the filter 91, 92 through the filter array eg. 96, 97. A filtered signal can be added to the filter feed signal before HRTF filter processing. This gives the reverberation a more plausible spatial-components and is likely to improve the .... experience.

The reverb generating filters 91, 92 may be a sparse tap FIR, an recursive algorithmic filter or a full convolutional FIR. In all these cases it may be beneficial to feed the outputs of the reverb back into the virtual speaker feeds. The result is likely to be most significant in the low resource system where a sparse tap FIR is used to simulate the reverb. Sparse tap reflection simulations then appear to emanate from sources outside of the listener rather than from the headphones.

Turning now to Fig. 8, there is shown a modified embodiment 100 similar to the embodiment 50 of Fig. 5. The arrangement includes the two sum and difference filters 101, 102 which are short time FAR approximation to the direct plus shadowed and the direct minus shadowed HRTF's four speakers located at around 30% either side of the list. However, the arrangement 100 of Fig. 8, an additional signal is derived as the sum 103 of the two inputs and fed to a single sparse cap reverberation FIR delay line 104. Two sparse tap outputs 105, 106 are derived from a set of coefficients within the FIR 104. This pair of signals 105, 106 is then added 107, 108 to the input stereo signals prior to the shuffling process 109. Thus the stereo sparse tap reverb is "binauralised".

The arrangement of Fig. 8 can be extended to a surround sound decoder somewhat to the arrangement of Fig. 6. Such an extension is illustrated in Fig. 9 with the portion 111

being similar to that of Fig. 6. The arrangement of Fig. 9 provides for the centre speaker feed 112 to be rendered as a virtual speaker panned midway between the front left and front right speakers. This is achieved by adding 113, 114 the centerfeed speaker 112 to the front left and front right speaker feeds. The rear speaker feeds 116, 117 have a separate shuffler 118 and some 119 and difference filter 120 to approximate the HRTF responses for speakers located 120° either side of the front of the listener. The outputs are then mixed together 122, 123 and fed into a single shuffler 124 so as to form the binaural outputs. Each of the inputs are summed 126 to form a single mono signal for reverb processing by a sparse tap reverb FIR filter 127. The reverb filter outputs are then added to the front speaker feeds 113, 114. Whilst further reverb signals could be added to the rear speaker feeds, it is generally advantageous for the system to throw images forward to overcome psycho-acoustic frontal .... confusion and elevation. Using only the front speaker positions for the reverb helps to throw the images forward and give a more convincing frontal sound.

Turning now to Fig. 10, in order to better describe the derivation of filter values for the sparse .... reverb FIR 127 of Fig. 9, a number of terms are defined. Firstly, the direct HRTF is defined as the transfer function from a virtual speaker location, 130, 131 to a persons ear 132 which is located on the same side of her head. The shadowed HRTF function is defined as the transfer function from the virtual speaker location eg. 130, 131 to the person's ear 133 on the opposite side of the head. An actual set of HRTF measurements can be used to approximate the filters. The frontal HRTFs can be measured from speakers located in front of the listener, 30° to each side. The rear HRTF can be measured from speakers located 120° to either side of the listener. Preferably, the HRTFs are equalized for maximum sound quality with good vocalisation properties.

The front sum filter 128 of Fig. 9 is an approximation of the sum and direct and shadowed frontal HRTF. The filter implementation can be a direct form transfer function (FIR) and (IIR) with a substantial FIR component allowing for non-minimum phase transfer function. The system orders can be selected by calculating a grid of approximation error versus FIR and IIR order. The Sum and Difference filters can be approximated with the order set at each point in the grid, then the error in the Direct and Shadowed HRTF plotted - this is shown in Fig. 11 and 12 for the front direct and shadowed response respectively. Prony analysis was used for the approximation. The plots exhibit "knee" characteristics demonstrating the significance of a certain order and diminishing returns beyond that. The order for the two frontal filters can be selected based on this information. Effective results were obtained with a FIR order of 14 and an IIR order of 4.

The front difference filter 129 can be an approximation of the frontal Direct HRTF minus the frontal Shadowed HRTF. The approximation can be carried out as described in the previous section resulting in an FIR order of 14 and IIR order of 4.

The rear sum filter 119 is an approximation of the rear Direct HRTF plus the rear Shadowed HRTF. The approximation can be carried out as described for the frontal filters. A FIR order of 25 and IIR order of 4 was selected.

The rear difference filter 120 is an approximation of the rear Direct HRTF minus the rear Shadowed HRTF. The approximation can be carried out as described for the frontal filters. A FIR order of 25 and IIR order of 4 was selected.

The reverb filter long delay line that is fed with a sum 126 of all the inputs (mono signal). Two sets of sparse tap coefficients are used to create two outputs from this delay line. The delay line 127 can be as long or as short as memory allows. A minimum length of around 300-400 taps

is preferred for reasonable results. The sparse tap coefficients are similar in properties but quite different in value. The actual taps used were generated by a random process with the following constraints:

* No taps are present in the first 300-400 taps. This is to create a gap between the initial HRTF response and the first early echoes. This is to prevent obscuring the spatial location in the initial HRTF.

* The taps decrease in amplitude with time. This is to model the attenuation of transmission through air and lossy reflection. The decrease is dithered. This level of detail is not necessary but for longer filters with many taps it produces much more natural sounding results.

* The taps increase in frequency with time. This is to model the increasing density of early echoes as the path length increases and the possible paths to the listener increases.

Several sets of random coefficients were created under these constraints and a set chose which looked to be evenly spread (not too clustered) and produced a good sound.

An example of the sparse tap filter is shown in Fig. 13.

Other methods and approximations for deriving the sparse tap coefficients may be sued but experimentation found this method to be most suitable.

The basic property of the reverb filter 127 is to create two uncorrelated outputs which contain information from the mono input signal dispersed in time without significant frequency coloration. Thus the filters could be recursive, reduced sample rate or involve other elaborate processing as memory and compute availability allows.

Fig. 14 and Fig. 15 respectively show the left and right outputs impulse from the reverb filter after passing through the frontal HRTFs. It can be seen that a significant amount of detail is obtained in the output filters for a relatively low amount of computation and

memory.

To facilitate discussion of important filter characteristics, some terminology is defined:

System: The system for virtual rendering of sources over headphones. In abstract form it consists of a device having a number of inputs (for each speaker position) and two outputs (for left and right ear of headphones).

Transfer Function:

The signal mapping from a given input to a given output. If a system has M inputs and N outputs there are MxN possible transfer functions. If the system is linear and time invariant then these transfer functions will be static and independent. These will often be referred to individually as Input to Output transfer function (for example Left to Left, Rear Left to Right).

Filter Characteristics

HRTFs

Each transfer function has an early part of the response which represents an approximation of a particular HRTF. This part will usually be up to 100 samples in length.

HRTF Symmetry

Where the input source virtual locations have some symmetry about the listener, the HRTFs may reflect this same symmetry. For example, where there are virtual speakers located 30° to the left and right of the listener, the HRFT or early part of the Left to Left transfer function would be identical to the early part of the Right to Right transfer function. So to the Left to Right and Right to Left would show similarity or equivalence in the early part.

Sparse Reverb

After the initial HRFTs a reverberant field approximation will be present in each transfer function. This approximation will be largely sparse. The properties of sparse are that the filter will be in some way degenerate, having identifiable degrees of freedom covering

a much smaller subset than that covered by complete freedom of the filter taps over the length of the filter.

The following are some possibilities for this sparse property:

* Actual sparse taps. The transfer function is predominantly zero with a number of non-zero taps. These are discrete and identical in all aspects other than amplitude and sign.

* Filtered sparse taps. The transfer function exhibits a repeated pattern at sparse positions in time. This is the result of passing a sparse tap type filter through a further filter to spread the taps. The sparse patterns will be identical in all aspects other than amplitude and sign. The patterns may overlap in which case it may not be so obvious to a casual observer of the presence of filtered sparse taps.

* Composite filtered sparse taps. Several unique sparse tap type sections may be created and passed through different filters. This will be identified by several different filter patterns being repeated in time identical in all aspect other than amplitude and sign. The filter patterns used by correspond to the early HRTFs of some or all of the systems transfer functions.

* Recursive sparse taps. As for the first point but with a recursive element. These sparse taps will continue indefinitely in time, decaying away as a geometric series.

* Recursive filtered sparse taps. The result of filtering a recursive sparse tap type implementation through specific filters and/or the HRTFs. This results in an algorithmic reverb with distinct filtered sparse taps initially, becoming an apparently complex response as time progresses. The filters may correspond to the early HRTFs of some or all of the systems transfer functions.

<u>Mono Reverb</u>

The reverberant part of the transfer functions can be derived from a mono or combined source. This is evidenced

by the equivalence of transfer functions from all inputs to a particular output. For example in the stereo virtual speaker example, the Left to Left and Right to Left transfer functions would exhibit very similar characteristics in the later part of the response. Any difference int he response could be attributable to a shift in time, scaling or simple filtering operation.

It would be further appreciated by a person skilled in the art that numerous variations and/or modifications any be made to the present invention as shown in the specific embodiment without departing from the spirit or scope of the invention as broadly described. The present embodiment is, therefore, to be considered in all respects to be illustrative and not restrictive.

We Claim:

1. An apparatus for creating, utilizing a pair of oppositely opposed headphones, the sensation of a sound source being spatially distant from the area between said pair of headphones, said apparatus comprising:

(a) a series of audio inputs representing audio signals being projected from an idealized speaker located at a spatial location relative to an idealized listener;

(b) a first mixing matrix means interconnected to said audio inputs for outputting a predetermined combination of said audio inputs as intermediate output signals;

(c) a filter system for filtering said intermediate output signals and outputting filtered intermediate output signals; said filter system including separate filters for filtering the direct response and short time response and an approximation to the reverberent response; and

(d) a second mixing matrix means combining said filtered intermediate output signals to produce left and right channel stereo outputs.

2. An apparatus as claimed in claim 1 wherein said first mixing matrix means outputs a linear combination of said audio inputs.

3. An apparatus as claimed in claim 1 wherein said first matrix means applies a time varying gain to said audio inputs.

4. An apparatus as claimed in any previous claim wherein said filters are independent of one another.

5. An apparatus as claimed in any previous claim wherein said audio inputs comprise Dolby AC-3 inputs.

6. An apparatus as claimed in any previous claim 1 to 4 wherein said audio inputs comprise stereo inputs.

7. An audio processing method for converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the

inputs so as to create the appearance of sound located around a listener, said method comprising:

filtering each of the Dolby AC-3 inputs utilising first filters constructed to simulate the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear;

applying a second filter to each of said inputs to simulate the reverberant tail of a suitably arranged virtual speaker to a corresponding listener's ear; and

adding together the outputs from said filtering step and said applying step to produce left and right stereo headphone outputs.

8. A method as claimed in claim 7 wherein said inputs are summed before being input to said second filters.

9. A method as claimed in claim 7 wherein said first filters comprise short filter lengths whereas said second filters comprise substantially longer filter lengths.

10. A method as claimed in claim 9 wherein said first filters are about 2,000 taps in length and said second filters are about 32,000 taps in length.

11. An audio processing apparatus for converting Dolby AC-3 inputs to stereo headphone outputs so as to substantially preserve the spatial components present in the inputs so as to create the appearance of sound located around a listener, said apparatus comprising:

a first series of early response filters for filtering said inputs so as to produce outputs simulating the early part of the response from a suitably arranged virtual speaker to a corresponding listener's ear;

a second series of reverberant tail filters for filtering said inputs so as to produce outputs simulating the reverberant tail response from a suitably arranged virtual speaker to a corresponding listener's ear; and

a left and right output combining means for combining the outputs of said first and second series of filters so as to produce left and right headphone outputs.

12. An audio processing apparatus as claimed in claim 11 wherein the number of reverberant tail filters is two and said inputs are summed together before input to said reverberant tail filters.

13. A method of processing stereo input sound sources for playback over headphones so as to create the sensation of sound originating from around a headphone listener, said method comprising the steps of:

(a) producing sum and difference signals from said stereo input sound sources;

(b) applying a direct ear response and shadow ear response filter to said difference signal to form a filtered difference output;

(c) applying a direct ear response, a shadow ear response and a reverberant response filter to said sum signal to form a filtered sum output;

(d) forming a first headphone output from the addition of said filtered difference output and said filtered sum output; and

(e) forming a second headphone output from the subtraction of said filtered difference output and said filtered sum output.

14. A method as claimed in claim 13 wherein said responses simulate head related transfer functions for the placement of virtual speakers at substantially 30 degrees to the horizontal plane.

15. A method as claimed in claim 13 wherein said filters comprise forming the following outputs:

$$Sum' = \left(\sqrt{(1-\alpha^2)}(D+S)+\alpha R\right) \otimes Sum$$

$$Diff' = \left(\sqrt{(1-\alpha^2)}(D-S)\right) \otimes Diff$$

where:

Sum and Diff are the sum signal and difference signal respectively;

Sum′ and Diff′ are the filtered sum output and filtered

difference output respectively;

D is the direct ear response - normalised to unity energy;

S is the shadowed ear response - scaled in proportion to D;

R is the reverberant response - normalised to unity energy;

α is the presence - the amount of reverberant feed in the mix.

16. A method as claimed in claim 13 wherein in said shadow ear response filter comprises a short FIR filter matching the frequency response and group delay of a signal derived from deconvolving a direct ear response from an appropriate shadowed response.

17. A method as claimed in claim 13 wherein said reverberant response filter approximates a delay line of between 5 - 10 ms

18. A method of processing Dolby AC-3 input sound sources for playback over headphones so as to create the sensation of sound originating from around a headphone listener, said method comprising the steps of:

(a) producing sum and difference signals from the Right Rear and Left Rear input signals;

(b) producing an intermediate front left signal from the addition of the front left signal and the center right signal;

(c) producing an intermediate front right signal from the addition of the front right signal and the center signal;

(d) applying separate HRTF signals to said intermediate signals;

(e) applying an anti-phase HRTF to said sum and difference signals;

(f) summing the outputs of steps (d) and (e) to produce left and right channels headphone signals.

19. A method as claimed in claim 18 wherein said

intermediate signals are delayed before the application of said HRTFs.

20. An apparatus for creating, utilizing a pair of oppositely opposed headphones, the sensation of a sound source being spatially distant from the area between said pair of headphones, said apparatus comprising:

(a) a series of audio inputs representing audio signals being projected from an idealized sound source located at a spatial location relative to the idealised listener;

(b) a first mixing matrix means interconnected to said audio inputs and a series of feedback inputs for outputting a predetermined combination of said audio inputs as intermediate output signals;

(c) a filter system of filtering said intermediate output signals and outputting filtered intermediate output signals and said series of feedback inputs, said filter system including separate filters for filtering the direct response and short time response and an approximation to the reverberant response, in addition to feedback reponse filtering for producing said feedback inputs; and

(d) a second matrix mixing means combining said filtered intermediate output signals to produce left and right channel stereo outputs.

21. An apparatus as claimed in claim 20 wherein a predetermined number of said feedback inputs are also input to said second matrix mixing means.

22. An apparatus as claimed in any previous claim wherein said feedback response filtering comprises a reverberation filter.

23. An apparatus as claimed in claim 22 wherein said reverberation filter comprises one of a sparse tap FIR, a recursive algorithmic filter or a full convolution FIR filter.

24. An apparatus as claimed in any of claims 20 to 23 wherein said audio inputs comprise a surround sound set of

signals.

25. An apparatus as claimed in claim 24 wherein said feedback inputs are mixed with the frontal portions of said audio inputs only.

26. An apparatus as claimed in any previous claim wherein said filter system includes a front sum filter filtering a summation of said audio inputs positioned in front of said idealized listener and said front sum filter comprises substantially an approximation of the sum of a direct and shadowed head related transfer function for said front inputs.

27. An apparatus as claimed in any previous claim 20 to 26 wherein said filter system includes a front difference filter filtering a difference of said audio inputs positioned in front of said idealized listener and said front difference filter comprises substantially an approximation of the difference of a direct and shadowed head related transfer function for said front inputs.

28. An apparatus as claimed in any previous claim 20 to 27 wherein said filter system includes a rear sum filter filtering a summation of said audio inputs positioned in rear of said idealized listener and said rear sum filter comprises substantially an approximation of the sum of a direct and shadowed head related transfer function for said rear inputs.

29. An apparatus as claimed in any previous claim 20 to 27 wherein said filter system includes a rear difference filter filtering a difference of said audio inputs positioned in rear of said idealized listener and said rear difference filter comprises substantially an approximation of the difference of a direct and shadowed head related transfer function for said rear inputs.

30. An apparatus as claimed in any previous claim 20 to 27 wherein said filter system includes a reverberation filter interconnected to the sum of said audio inputs.

Dated this 31st day of March 1998

Lake DSP Pty. Ltd.
By their Patent Attorneys
GRIFFITH HACK

Abstract

An apparatus for creating, utilizing a pair of oppositely opposed headphones, the sensation of a sound source being spatially distant from the area between the pair of headphones is disclosed, the apparatus comprising: (a) a series of audio inputs representing audio signals being projected from an idealized sound source located at a spatial location relative to the idealised listener; (b) a first mixing matrix means interconnected to the audio inputs and a series of feedback inputs for outputting a predetermined combination of the audio inputs as intermediate output signals; (c) a filter system of filtering the intermediate output signals and outputting filtered intermediate output signals and the series of feedback inputs, the filter system including separate filters for filtering the direct response and short time response and an approximation to the reverberant response, in addition to feedback response filtering for producing the feedback inputs; and (d) a second matrix mixing means combining the filtered intermediate output signals to produce left and right channel stereo outputs. Preferably, a predetermined number of the feedback inputs are also input to the second matrix mixing means. The feedback response filtering can comprise a reverberation filter. The reverberation filter can comprise one of a sparse tap FIR, a recursive algorithmic filter or a full convolution FIR filter and the audio inputs can comprise a surround sound set of signals.

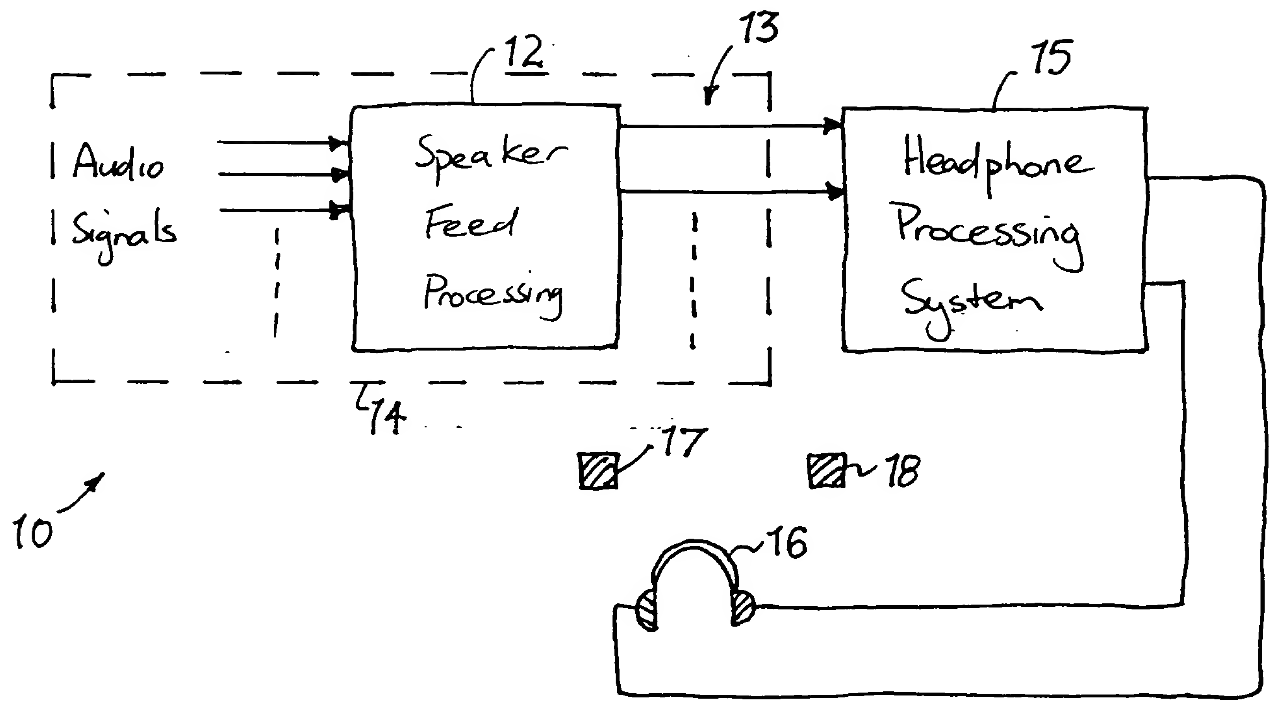

Audio
Signals
Speaker
Feed
Processing
12
13
Headphone
Processing
System
15
14
17
18
16
10

FIG. 1

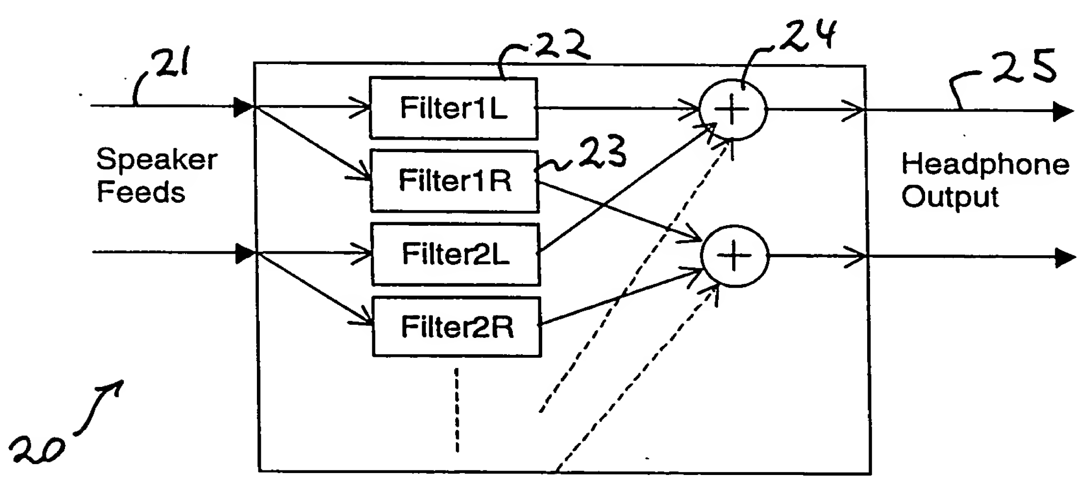

21
22
24
25
23
Speaker
Feeds
Filter1L
Filter1R
Filter2L
Filter2R
Headphone
Output
20

FIG. 2

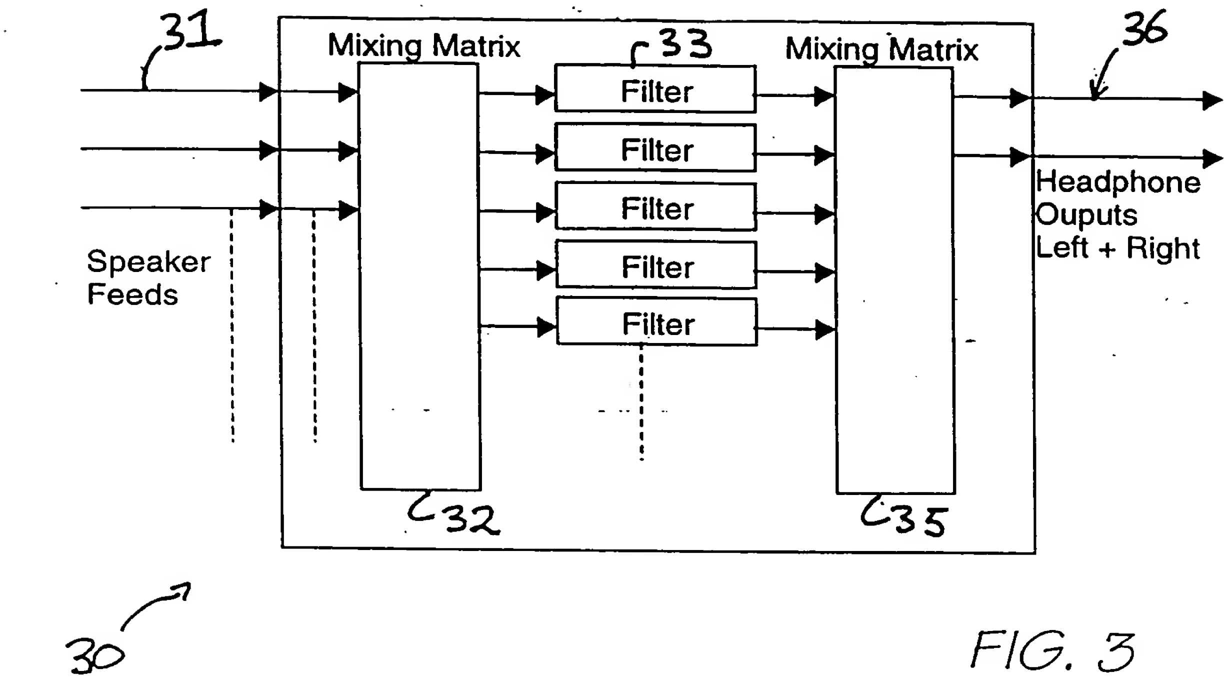

31
Mixing Matrix
33
Filter
Filter
Filter
Filter
Filter
Mixing Matrix
36
Headphone
Ouputs
Left + Right
Speaker
Feeds
32
35
30

FIG. 3

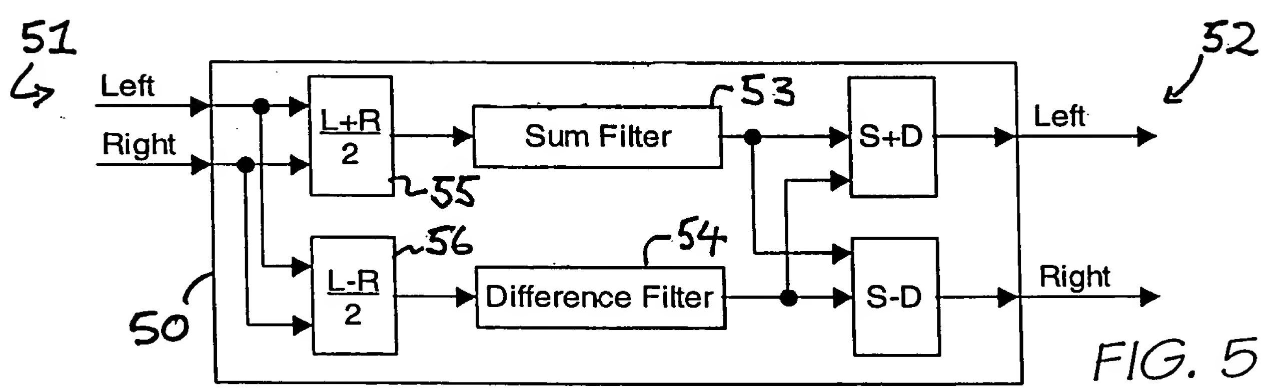

51
Left
Right
L+R
2
55
Sum Filter
53
S+D
52
Left
L-R
2
56
Difference Filter
54
S-D
Right
50

FIG. 5

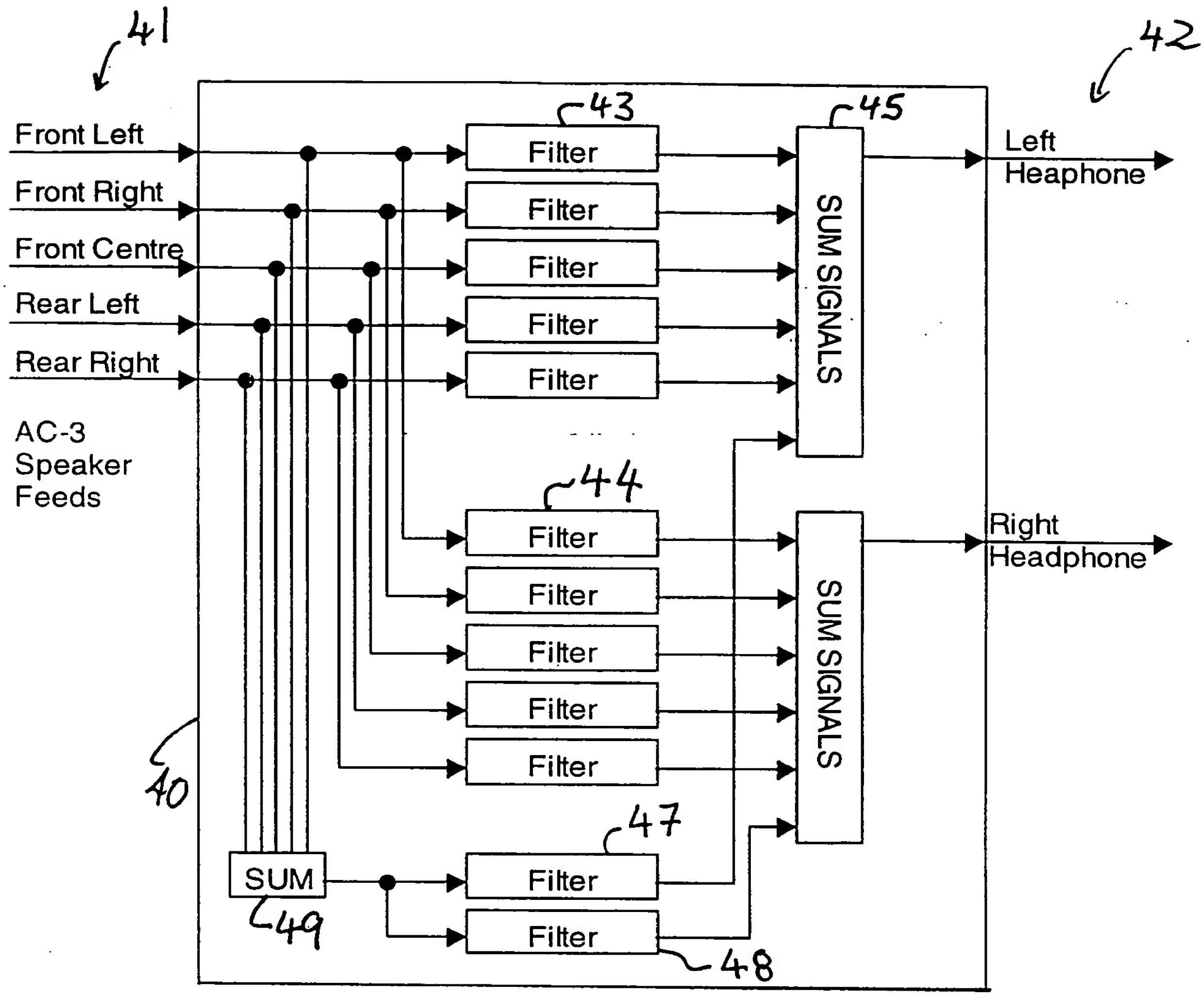

41
42
Front Left
Front Right
Front Centre
Rear Left
Rear Right
AC-3
Speaker
Feeds
43
44
45
Filter
Filter
Filter
Filter
Filter
Filter
Filter
Filter
Filter
Filter
Filter
Filter
SUM SIGNALS
SUM SIGNALS
Left
Heaphone
Right
Headphone
40
SUM
49
47
48

FIG. 4

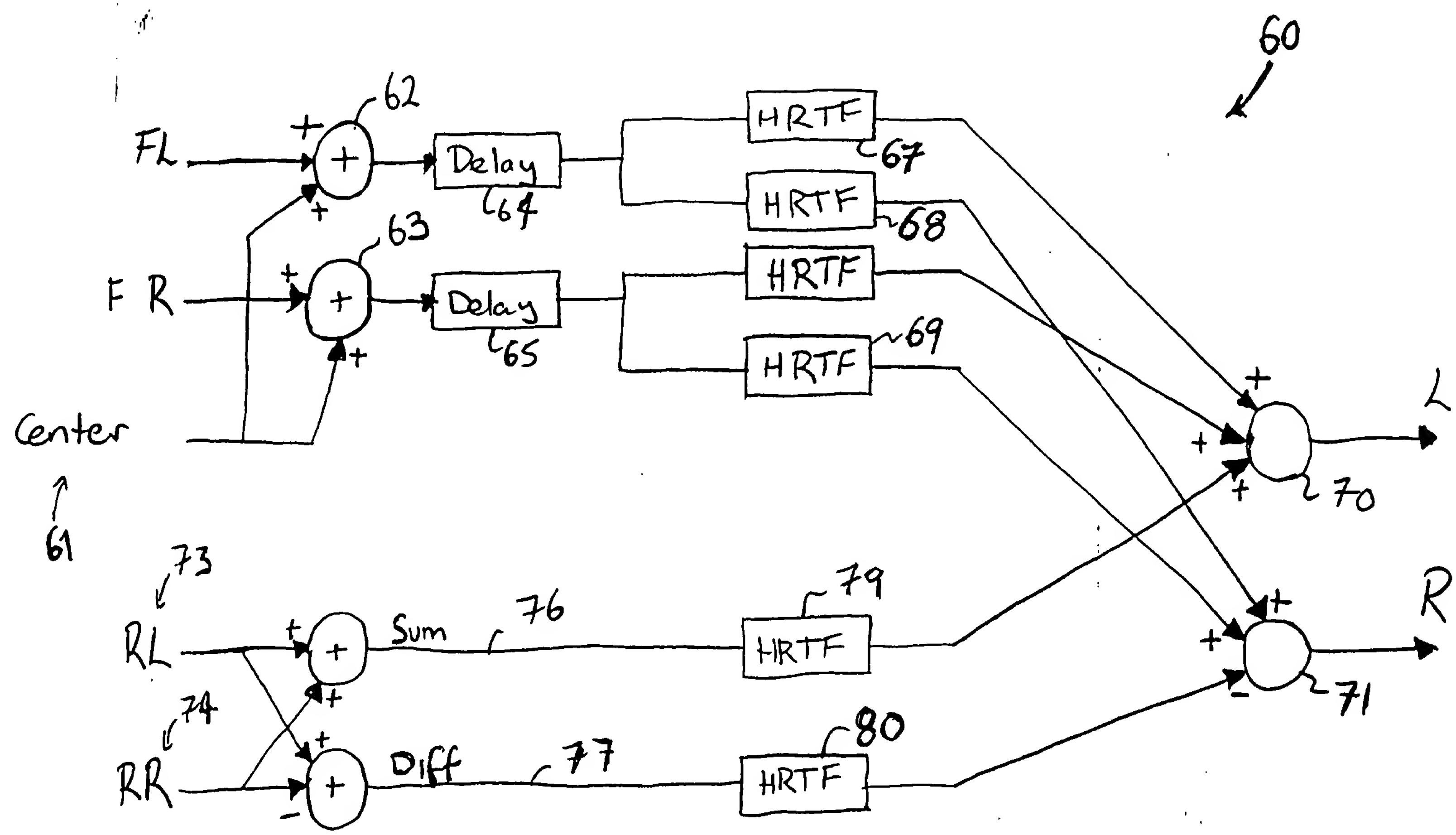
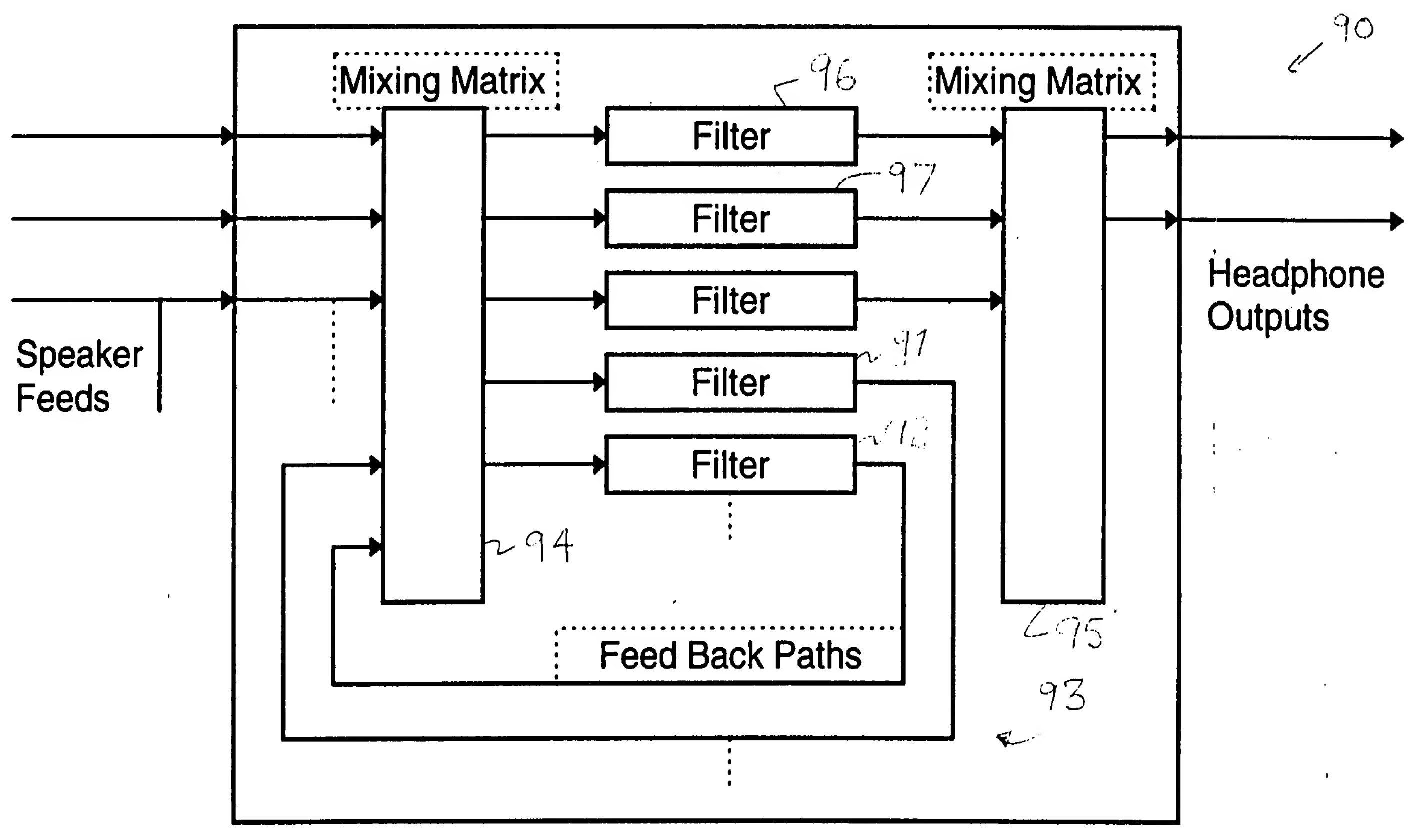

60
FL
FR
Center
61
62
63
Delay
64
Delay
65
HRTF
67
HRTF
68
HRTF
HRTF
69
RL
73
RR
74
Sum
76
Diff
77
HRTF
79
HRTF
80
70
71
L
R

FIG. 6



90
Mixing Matrix
Filter
Filter
Filter
Filter
Filter
96
97
91
92
94
Mixing Matrix
95
93
Feed Back Paths
Speaker
Feeds
Headphone
Outputs

FIG. 7

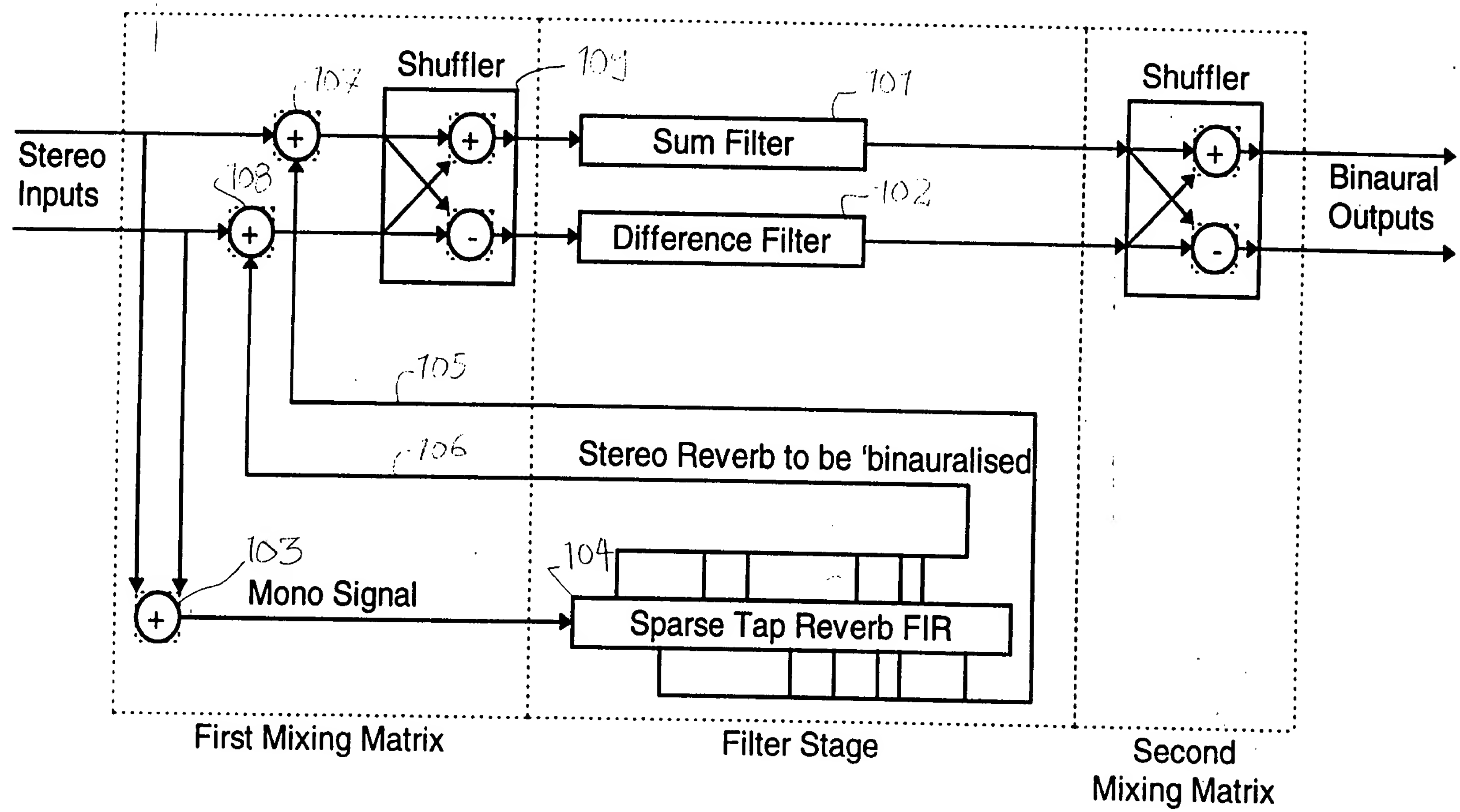

Stereo Inputs
107
108
Shuffler
109
Sum Filter
101
Difference Filter
102
Shuffler
Binaural Outputs
105
106
Stereo Reverb to be 'binauralised
103
Mono Signal
104
Sparse Tap Reverb FIR
First Mixing Matrix
Filter Stage
Second Mixing Matrix

FIG. 8

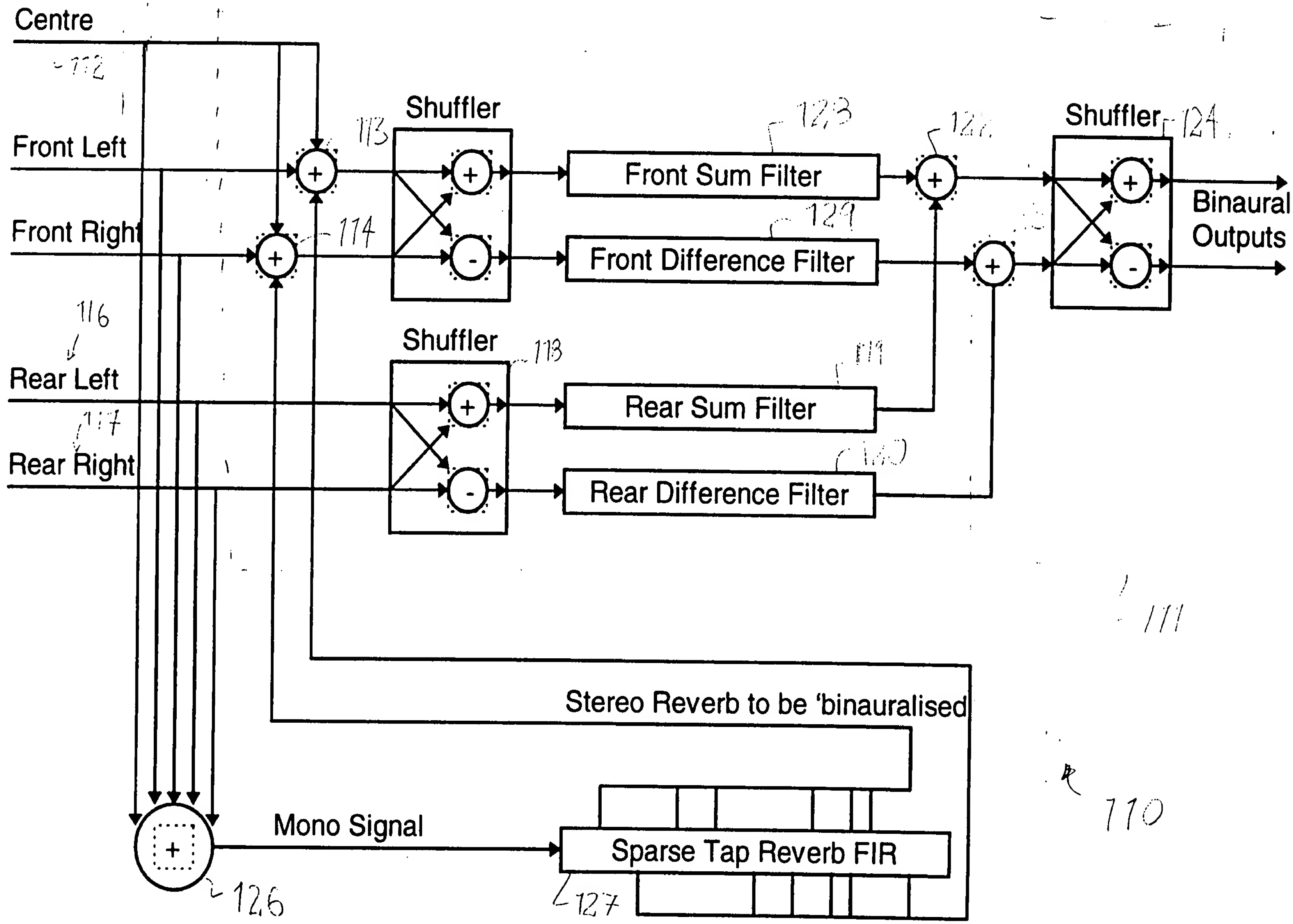

Centre
Front Left
Front Right
Rear Left
Rear Right
Shuffler
Shuffler
Front Sum Filter
Front Difference Filter
Rear Sum Filter
Rear Difference Filter
Shuffler
Binaural Outputs
Stereo Reverb to be 'binauralised
Mono Signal
Sparse Tap Reverb FIR
112
113
114
116
117
118
123
124
126
127
110
111

FIG. 9

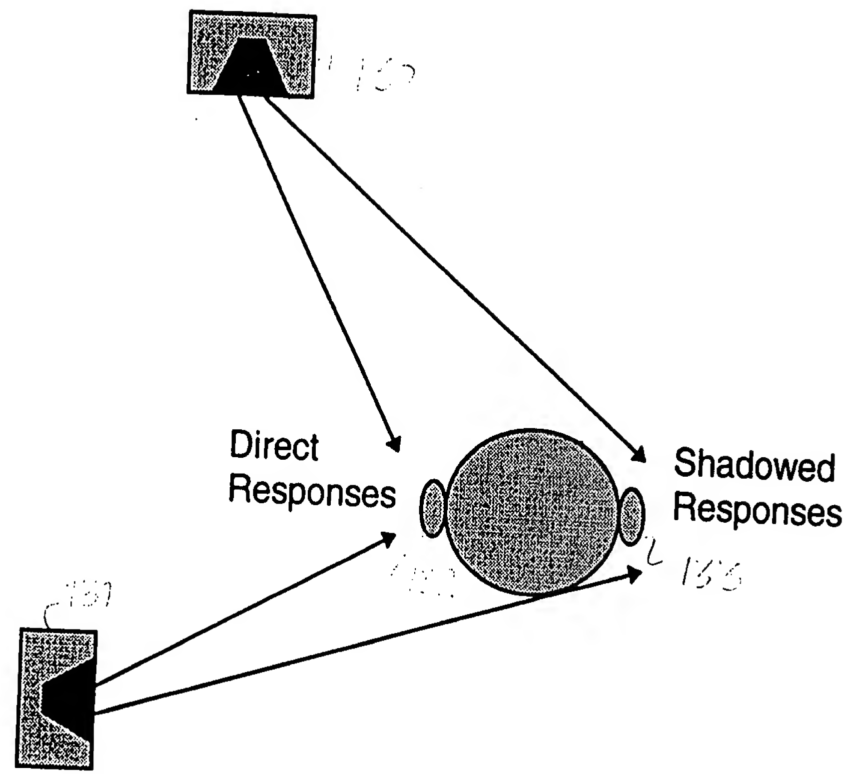


FIG. 10

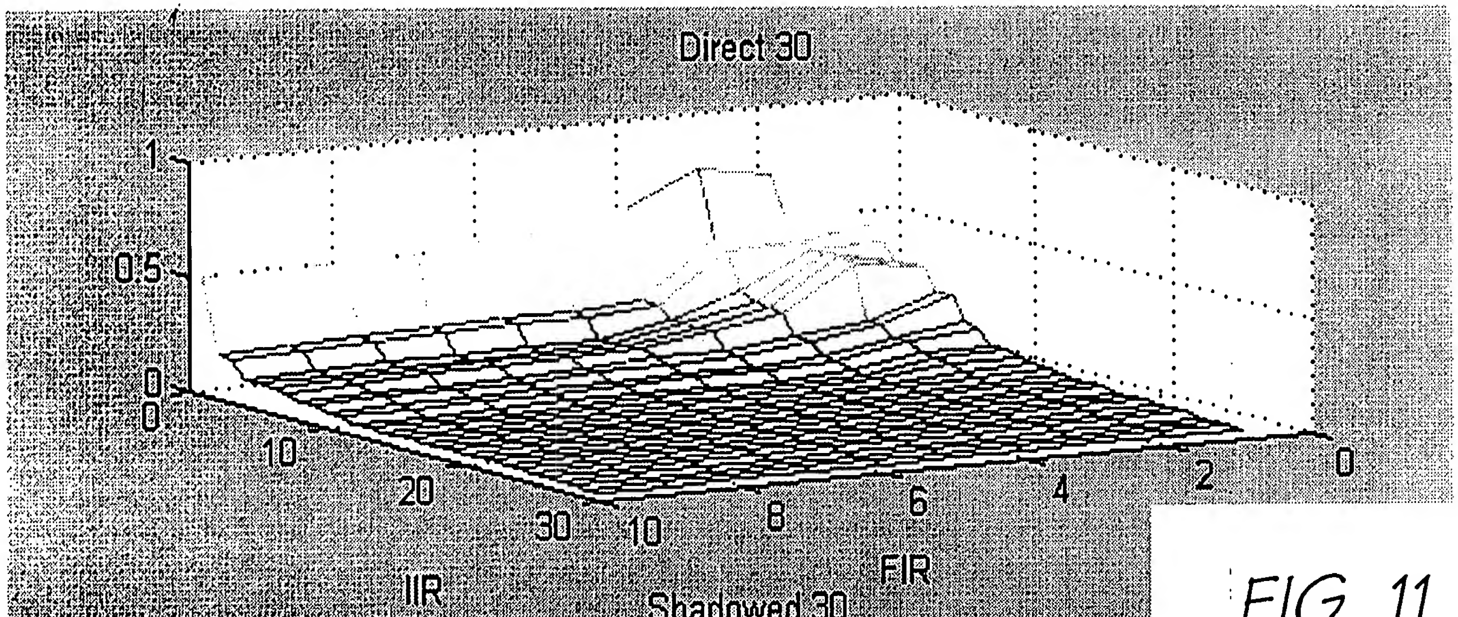

Direct 30
1
0.5
0
0
10
20
30
10
8
6
4
2
0
IIR
FIR

FIG. 11

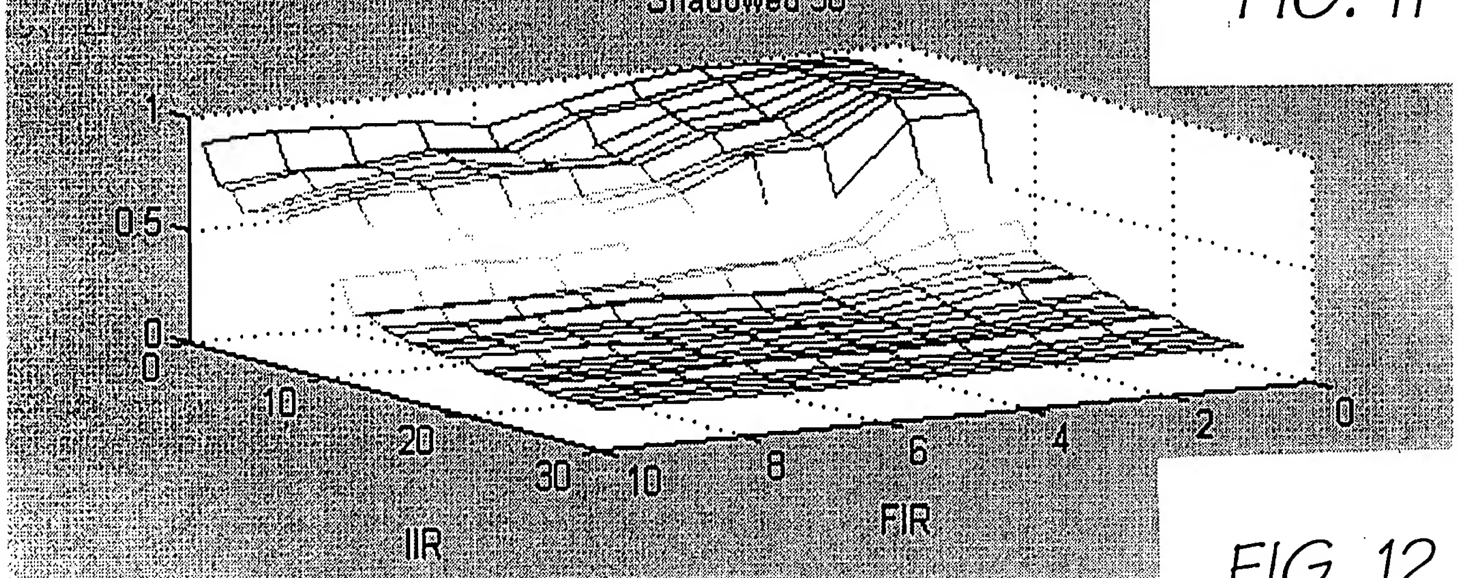

Shadowed 30
1
0.5
0
0
10
20
30
10
8
6
4
2
0
IIR
FIR

FIG. 12

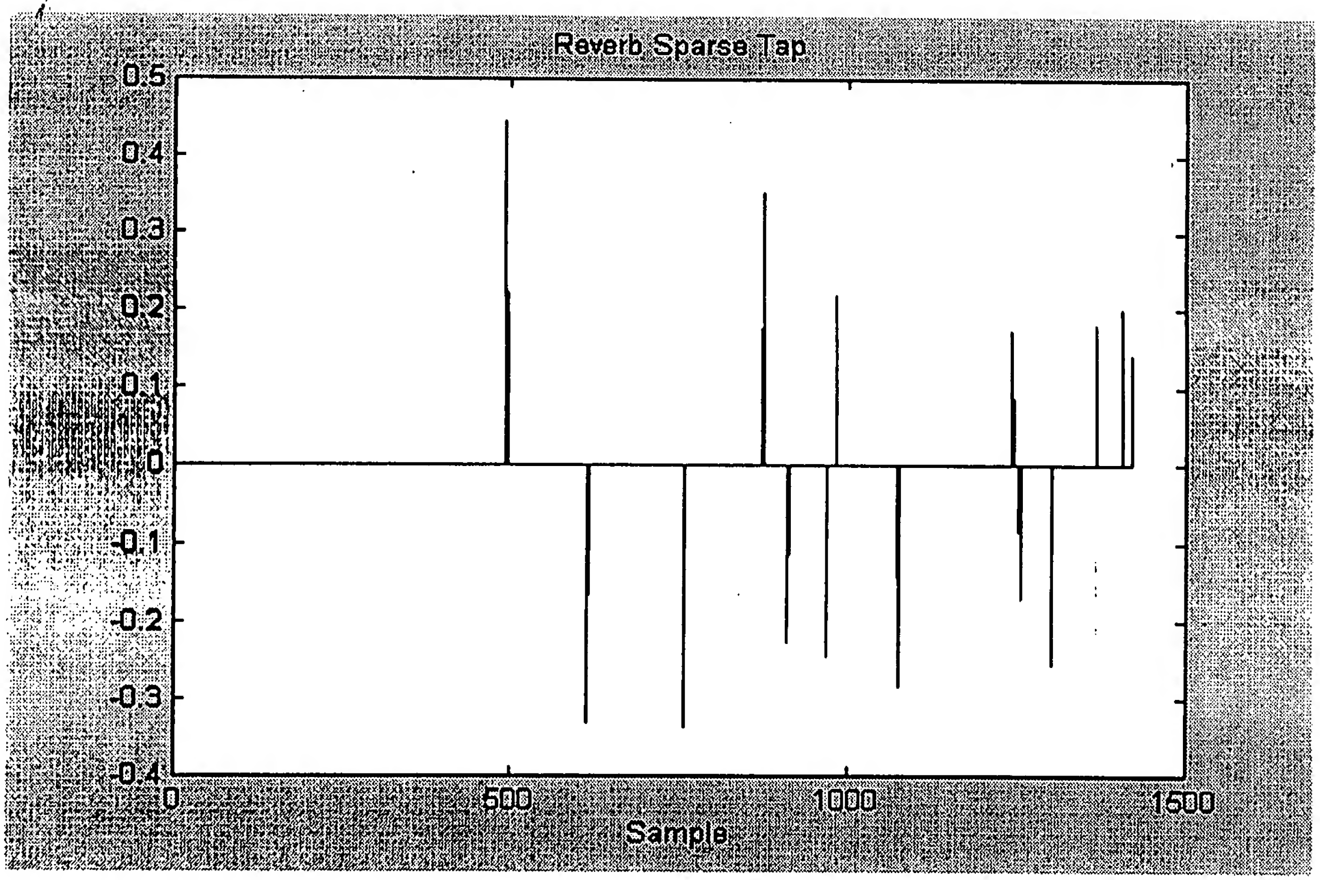

Reverb Sparse Tap
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
0
500
1000
1500
Sample

FIG. 13

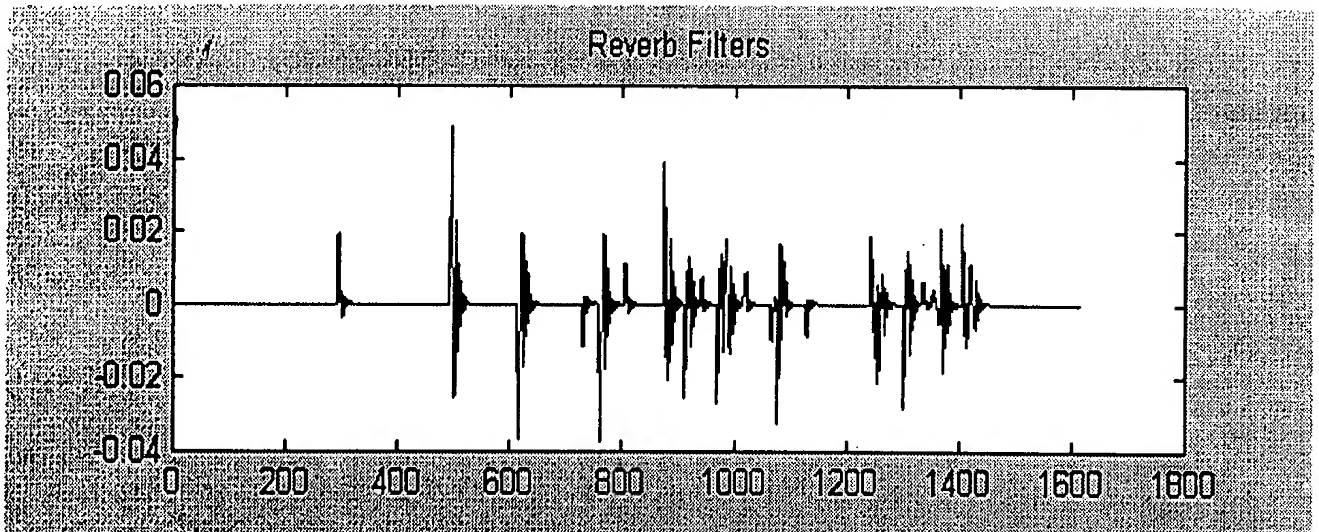

Reverb Filters
0.06
0.04
0.02
0
-0.02
-0.04
0
200
400
600
800
1000
1200
1400
1600
1800

FIG. 14

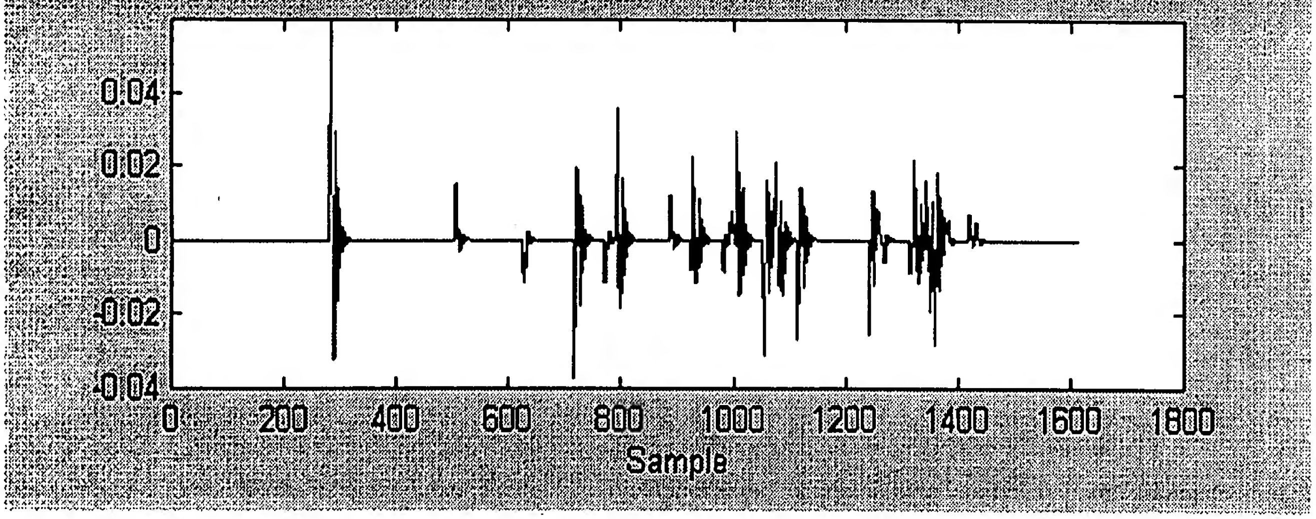

0.04
0.02
0
-0.02
-0.04
0
200
400
600
800
1000
1200
1400
1600
1800
Sample

FIG. 15